Neuroradiology (2003) 45: 691–699
DOI 10.1007/s00234-003-1026-9

DIAGNOSTIC NEURORADIOLOGY

M. Buerki
K.O. Lovblad
H. Oswald
A.C. Nirkko
P. Stein
C. Kiefer
G. Schroth

Multiresolution fuzzy clustering of functional MRI data

Received: 25 October 2002
Accepted: 22 April 2003
Published online: 27 August 2003

This research is supported by a grant from the Swiss National Science Foundation (grant no. 3100-66348.01)

M. Buerki · K.O. Lovblad · P. Stein
C. Kiefer · G. Schroth
Department of Neuroradiology,
DRNN, Inselspital, Bern, Switzerland

K.O. Lovblad (✉)
Unité de NeuroradiologieService de
Radiodiagnostic, Etage P,
Hôpital Cantonal Universitaire de
Genève, 24 rue Micheli-du-Crest,
1211 Genève 14, Switzerland
E-mail: Karl-Olof.Lovblad@hcuge.ch

H. Oswald
T-Systems, Bern, Switzerland

A.C. Nirkko
Department of Neurology, University
Hospital of Bern, 3010 Bern, Switzerland

Abstract Recent developments in the analysis of functional MRI data reveal a shift from hypothesis-driven statistical tests to unsupervised strategies. One of the most promising approaches is the fuzzy clustering algorithm (FCA), whose potential to detect activation patterns has already been demonstrated. But the FCA suffers from three drawbacks: first the computational complexity, second the higher sensitivity to noise and third the dependence on the random initialization. With the multiresolution approach presented here, these weak points are significantly improved, as is demonstrated in our tests with simulated and real functional MRI data.

Keywords fMRI · Multiresolution · Fuzzy clustering

Introduction

In functional MRI (fMRI), activation patterns are usually retrieved by voxel-wise testing for correlation of the measurements with a paradigm-defined input function (e.g. SPM [1]). In contrast to these standard analysis tools, the fuzzy clustering algorithm (FCA) [2] is a way of investigating complex data without imposing prior assumptions on the results that are searched for. It has been shown that the FCA can reliably detect activation patterns [3, 4], even under difficult conditions and in complex studies where several different functional components are involved [5].

The FCA-based analysis of fMRI time series considers each of the N voxel time courses as a P-dimensional vector $\mathbf{X}=(X_1,...,X_p)$, where N is the number of voxels making up the volume of one MR acquisition and P is the number of acquisitions. The algorithm aims to find C vectors $\mathbf{v}_1,...,\mathbf{v}_C$—the prototypes or centroids for C different clusters—such that the objective function

$$J_m(U,V)=\sum_{i=1}^{N}\sum_{j=1}^{C}\left(u_{ij}\right)^m d^2\left(\mathbf{x}_i,\mathbf{v}_j\right) \qquad (1)$$

is minimized. The scalars $u_{ij}\in[0,1]$ describe the degree of membership of the time course X_i to the jth cluster, where $u_{ij}=1$ means exclusive membership and $u_{ij}=0$ is no membership. d^2 is any inner product metric on $\mathbf{R}^P$, the real number $m>1$ regulates the fuzziness of the clustering and V represents the set of centroids $\mathbf{v}_1,...,\mathbf{v}_C$. Necessary conditions for minimizing J_m are:

$$u_{ik}=\frac{1}{\sum_{j=1}^{C}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}}\; 1\leqslant i\leqslant N, 1\leqslant j\leqslant C \tag{2}$$

where $d_{ik}=d(\mathbf{x}_i,\mathbf{v}_k)$ and

$$\mathbf{v}_j=\frac{\sum_{k=1}^{N}\left(u_{jk}\right)^m\mathbf{x}_k}{\sum_{k=1}^{N}\left(u_{jk}\right)^m}\; 1\leq j\leq C. \tag{3}$$

A local minimum is now found by iteratively computing the centroids $\mathbf{v}_1,...,\mathbf{v}_C$ and the membership matrix $U=[u_{ij}]$ with formulas (3) and (4) until the distance between two consecutively calculated membership matrices $U^{(k)}$ and $U^{(k+1)}$ is less than a defined threshold, i.e.

$$\Delta U=\left\|U^{(k+1)}-U^{(k)}\right\|\leq\varepsilon, \tag{4}$$

where the threshold is a small number, e.g. $\varepsilon=0.01$.

To start, the FCA needs initialization. Since we assume that nothing is known about the data fed into the FCA, the initialization must be a random guess. The initialization is usually done for the membership matrix U, but starting with some randomly chosen centroids would also work. Following Bezdek [2], U is usually initialized as

$$U=\left(1-\tfrac{\sqrt{2}}{2}\right)U_C+\tfrac{\sqrt{2}}{2}U_r \tag{5}$$

with $U_C=[1/C]$ and U_r is a random hard partition.

Limitations of the FCA

The analysis of fMRI data with the FCA suffers from three main drawbacks: first the solution and speed of convergence depend strongly on the (requisite) random initialization of the algorithm, second the FCA is fairly sensitive to noise, and third the high computational complexity renders the FCA very time-consuming. This section discusses approaches proposed by others to deal with these problems.

fMRI studies typically consist of 50 or more acquisitions. Supposing a volume of 64×64×64 voxels per scan leads to vectors of dimension $P=50$ or more and a membership matrix U of dimension C×262,144, resulting in an enormous slowdown of the computationally very expensive FCA. This huge number of time courses can be reduced by discarding the background voxels in fMRI datasets, since they certainly do not contribute anything to brain activation patterns. This can be effected by a simple mean thresholding, i.e. all time courses with a mean value less than a specified threshold are eliminated. This is a standard method also applied in hypothesis-driven analysis procedures to both improve the statistical power and to save time by reducing the number of statistical comparisons. Eliminating the background voxels reduces the volume of a scan to approximately 25–30% of its original size. Another efficient way to speed up the FCA is the preselection (screening) of potentially interesting time courses, such that those time courses where only noise is expected are discarded [6]. There exist several preselection methods, such as spectral peaks, autocorrelation and novelty indices. Of course, mean thresholding is also a kind of preselection. As a side effect, the reduction of the data to some potentially interesting subset renders the algorithm more robust, but the sensitivity to noise remains an issue. The noise problem should rather be approached via the fuzzy index m in formula (1). But unfortunately, there is no theoretical basis currently known for an optimal choice for the value of m and, consequently, only empirical solutions, such as ROC methods, are available. Although fairly successful, preselecting data by some criterion is a step back towards model-driven analysis and should therefore be used with caution. Furthermore, screening methods such as spectral peaks have lower sensitivity when the time courses of interest are not periodic, because the power spectral density has much less pronounced peaks that would indicate the presence of a signal. But the restriction to periodic or nearly periodic signals is rather severe with regard to the increasingly complex paradigms developed in fMRI and event-related fMRI studies.

The random initialization has a considerable effect on the convergence of the algorithm in terms of both speed and solution, as can be seen in the Results. But the concept of the random starting point is intrinsic to the FCA and thus remains a problem.

Multiresolution FCA

In [7], we introduced the multiresolution FCA (MFCA) to reduce the computational complexity. The MFCA takes advantage of the fact that the vectors under investigation derived from fMRI data have a spatial structure in terms of neighbourhoods. Two neighbouring vectors are likely to be more or less correlated, allowing for a multiresolution approach, that first scales down the data volumes. Starting with the lowest resolution, the FCA is applied to that level and then the computed centroids are used as initial values for the FCA for the next-highest level of resolution, and so on until the original resolution is reached. Since the processing of all lower resolution levels is virtually an initialization of the FCA at the full resolution level (in fact a very good initialization), it will converge quite rapidly to a local minimum.

As we shall show here, the MFCA not only speeds up the algorithm, it improves precisely those weak points of the standard FCA mentioned in the previous section. First, it depends much less on the random initialization and is therefore more robust. Second, the downsampling of the data volume has a smoothing effect that renders the MFCA less sensitive to noise. And third, as already mentioned, the computationally very complex algorithm becomes faster, since each level of resolution reduces the amount of data from its preceding level by a factor of 2 per dimension. A three-dimensional data set is therefore reduced to 1/8 or 12.5% per level, so that the FCA computation becomes extremely fast at lower levels. Another advantageous property of the MFCA is that it can be combined easily with other methods such as preselection that aim to optimize the algorithm. Joining the MFCA with preselection (mean thresholding, spectral peaks, autocorrelation and combinations of these three) combines the advantages of both approaches and leads to a fast and robust clustering of the data.

To further improve the performance of the algorithm, we vary the stopping criterion for the different levels. To justify this step, recall that the processing of the low resolution levels of the MFCA serves as an initialization of the FCA at the full resolution level and is already fairly close to the final solution. Therefore, the risk of starting at a "bad place" that would lead us to a weak local minimum is significantly smaller and, more important for our case, the starting point will not lie at a location that is far from a minimum of $\boldsymbol{J}_m$ and at the same time has a small derivative $D\boldsymbol{J}_m$ (which with random initialization of the FCA can happen quite frequently). Consequently, we can relax the stopping criterion for the algorithm, e.g. from $\varepsilon = 0.01$ to $\varepsilon = 1$, resulting in a significantly improved performance without loss of quality.

It is clear that the choice of ε should depend on the chosen matrix norm. For example, taking the maximum norm

$$\|U\|_{\max} = \max_{i,k}\{u_{ik}\}$$

requires a smaller ε than the sum-squared difference

$$\|U\|_{ssqd} = \sum_i \sum_k u_{ik}^2$$

to achieve (ideally) the same results. But, when using the latter, it should be noted that the matrix distance depends strongly on the dimensions of U, i.e. the number of time courses N and the number of clusters C. This applies especially to the case of the MFCA, where N differs from level to level. Obviously, ε should decrease with the dimensions of U. This would justify the use of a gradually increasing ε. On the other hand, taking the same ε for all levels results in a stronger stopping criterion at the high resolution levels. We chose an intermediate way by adapting the ε only at the final level, keeping in mind that the processing of the low resolutions is fast. Our tests with simulated and real fMRI data showed that the MFCA with $\varepsilon = 0.01$ at the levels $L-1$ to 1 and $\varepsilon = 1$ at level 0 worked perfectly well, whereas the single-level FCA failed in about 40% of cases when setting $\varepsilon = 1$.

It is important to check the number of initial clusters C against the number of time courses N in the level of lowest resolution to avoid the case $N < C$.

Another issue of interest is the kind of filter used to down-sample the volumes. There is an extensive amount of literature addressing this point and our comparisons of several filters have found that the simplest (and fastest), on the Haar wavelet based down-sampling operator is quite comparable regarding the robustness and convergence of the MFCA [8].

Materials and methods

We have tested our algorithm on simulated datasets as well as on real fMRI data. Each dataset was classified into $C = 4$ clusters with the standard FCA (one level) and with the multi-resolution FCA with two, three and four levels. The stopping condition of the FCA must be clearly defined in order to prevent the algorithm getting stuck at a bad minimum. However, level 1 is already very close to the solution so the risk of landing at a bad minimum no longer exists. This does not complicate performance measurements. Indeed we only count the number of necessary iterations in order to achieve the same (or better) result.

To compare the performance of our algorithm with the standard FCA, we have chosen the following machine-independent measure: we counted the total number of iterations I_L needed by the MFCA operating on L levels until the criterion for convergence was reached by weighting the number of iterations for each level using the formula

$$I_L = \sum_{l=0}^{L-1} \frac{\bar{t}_l}{\bar{t}_0} \cdot i_l, \tag{6}$$

where $\bar{t}_i$ is the mean value of the time consumed for one iteration and i_l is the number of iterations at level l. The levels are always labelled from 0 (the full resolution) to the number of levels L minus one (the least resolution).

The comparison was then expressed as a speed-up factor by calculating $sf = I_1/I_L$, i.e. by dividing the number of iterations in the standard FCA case by the total number of iterations needed by the L-level MFCA.

We generated a three-dimensional dataset of volume 64×64×32 voxels with 50 "scans", including a lower and a higher background level with a value of 30 and 22 respectively. Onto the higher level, we put two spatially separated "signals" over time, one describing a single peak with slow decay and the other as a periodic box car function. The spatial extent of these insertions was 16×16×8 voxels. Finally, we added gaussian noise in two different contrast to noise ratios, $CNR_1 = 1$ and $CNR_2 = 2$. Each of these two datasets has been analysed with various combinations of preprocessing and preselection steps. Throughout this work, the preprocessing steps are abbreviated with lower-case letters (m, mean subtraction; n, normalization; d, detrend) and the preselections with capital letters

(*S*, spectral peaks; *A*, autocorrelation; *M*, mean thresholding), and 0 (zero) indicates that no option was selected at a particular step. Each combination is abbreviated to a preprocessing code of the form preprocessing_preselection. The preprocessing applied to the time courses has a larger effect on the convergence of the algorithm than the preselection because preprocessing changes the distances between the vectors [9] whereas preselection mainly reduces the number of vectors to classify.

Testing the noise sensitivity

To test the noise sensitivity, we evaluated the datasets under various preprocessing options with the FCA and the two-, three- and four-level MFCA (labelled 1*L*, 2*L*, 3*L* and 4*L* respectively). The options were 0_0, 0_S, 0_A, m_0, m_S, m_A, n_0, n_S, n_A, and each run was repeated 30 times.

Testing with real fMRI data

To evaluate the MFCA with real fMRI data, we took a dataset of size 128×128×8 voxels with 48 scans. The underlying paradigm was a language task arranged in a standard block design. The volumes have not been corrected for head motion. For preprocessing, we used mean subtraction. This step was applied after preselection to allow for mean thresholding. Therefore, we would expect four different types of time courses in the dataset: the task-related signal, two opponent head-motion-correlated "signals" and noise. To see that head motion will result in two classes, suppose that the subject moved his head towards the left. Some pixels located on the left of the head will then move into the head and produce a rising signal whilst some pixels at the right side of the head will move outside the head and produce a descending signal.

We tested the MFCA under all possible combinations of the preselection options *M*, *S* and *A*, leading in total to eight tests that were repeated 30 times each. We let the algorithm search for $C=4$ clusters.

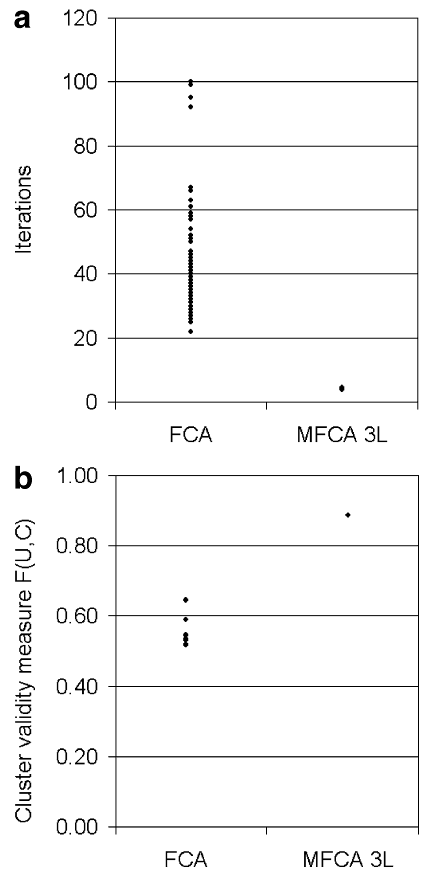


Fig. 1a, b Convergence of the standard FCA compared with the three-level MFCA (100 repetitions with random initialization). **a** Number of iterations until convergence. **b** Cluster validity measure of found solutions

Results

Dependence on the random initialization

The convergence of the MFCA is very robust compared with the FCA. This is illustrated in Fig. 1. The dataset with $CNR_1=1$ was evaluated with the FCA and the three-level MFCA 100 times each under the same conditions (preprocessing option 0_S). The number of iterations needed by the FCA varies over a broad range, indicating a strong dependence on the initialization. Following a common convention we stopped the computation when it did not converge within a reasonable time, i.e. we set the maximal number of iterations to 100. The MFCA, on the other hand, is extremely stable, showing almost no variability in the number of iterations required. The achieved speed-up for this case is up to 24.2, with an average of 11.4.

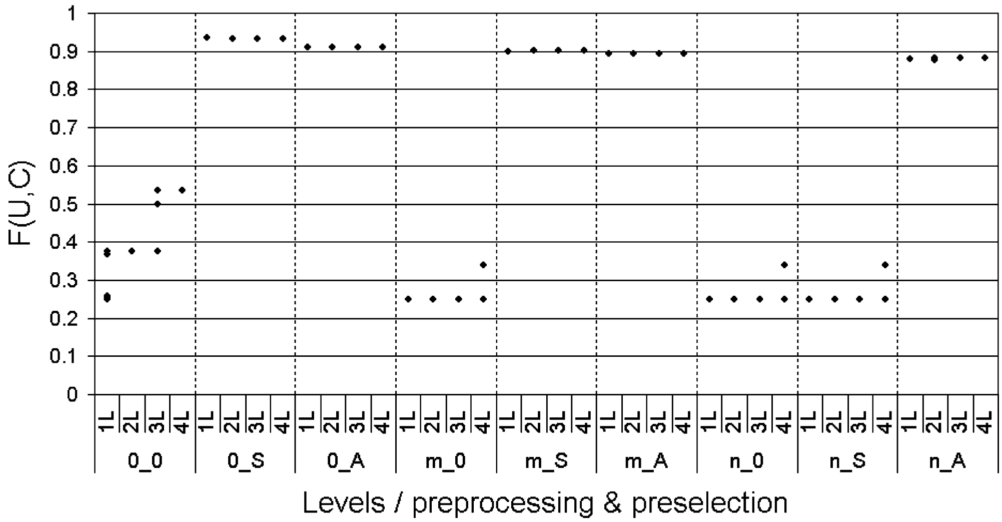
For many applications, speed might not matter. But, more importantly, the FCA did not always converge to the same solution. One way to measure the quality of a solution is a cluster validity measure [2]. There exist numerous such measures. We used the so-called partition coefficient $F(U,C)$, which is defined as follows:

$$F(U,C)=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}^2 \tag{7}$$

where again C is the number of clusters and N is the number of time courses.

The partition coefficient is a number in the closed interval $[1/C,\ 1]$. Low values correspond to large overlaps of different clusters and a value of 1 would mean that all clusters are strictly disjoint. A value of $1/C$ would stand for the case where all clusters are identical. Thus, we can see that the FCA converged to different solutions, depending on the initial conditions. These different solutions can be divided into three groups: the first group, with a $F(U,C)$ of 65%, has a relative group size of 34%; the second group, with a $F(U,C)$ of 0.59 has a 56% relative group size; and the third group, with a $F(U,C)$ between 0.51 and 0.55, has a relative group size of 10%. All runs in group 3 have not met the convergence criterion after 100 iterations. Note that the cluster validities are all below 0.70 because the algorithm was not able to extract all different classes of signals that

Fig. 2 Cluster validity measures $F(U,C)$ for the dataset with CNR = 2 and C = 4. For the distribution of the different solutions under a given condition see Table 1



exist in the dataset and at least two clusters were identical. In contrast to this, the MFCA could discriminate all clusters in all the test cases. Accordingly all values are about 0.89, indicating that the algorithm always converged to the same solution.

Noise sensitivity

The noise sensitivity results are reported in Fig. 2 and Table 1. For the dataset with CNR = 2, the FCA and the MFCA yielded the same results for the options 0_S, 0_A, m_S, m_A and n_A. The cluster validity measures achieved by the MFCA are slightly lower than for the FCA in these cases. This is because we set the threshold for the matrix distance in formula (4) to $\varepsilon = 1$at the full resolution level instead of $\varepsilon = 0.01$, forcing the algorithm to stop earlier. But the values differ less than 0.1% and have no measurable effect on the found centroids.

The values of $F(U,C)$ are all in the range from 0.88 to 0.93, which indicates a very good discrimination of the four classes present in the dataset. For option 0_0 (no preprocessing and no preselection), the results illustrate the power of the MFCA. The standard FCA converged in 27% of the runs to the worst possible solution with all four centroids being identical, i.e. $F(U,C=4) = 0.25$. In the other 73%, the FCA was only able to find the two background levels, and could not detect any signal. The MFCA on the other hand yielded better solutions the more levels were applied. The two-level MFCA could in all cases find the two background levels, the three-level MFCA found in some cases even the periodic boxcar signal and the four-level MFCA converged in all cases to the solution with three distinct centroids. Apparently, options m_0, n_0 and n_S were far more difficult to evaluate. The FCA as well as the two- and three-level MFCA ended up with the worst solution. Only the four-level MFCA found at least the two background levels.

Table 1 Distribution of the different solutions under the same preprocessing and preselection conditions for a given depth of the MFCA (number of levels) for CNR = 2

Preprocessing/preselection Number of levels	Cluster validity measures $F(U,C)$ and their relative frequency		
0_0 1L	0.25 (27%)	0.37 (73%)	
0_0 3L	0.37 (77%)	0.50 (13%)	0.54 (10%)
m_0 4L	0.25 (3%)	0.34 (97%)	
n_0 4L	0.25 (40%)	0.34 (60%)	
n_S 4L	0.25 (63%)	0.34 (37%)	

For the dataset with CNR = 1, the resulting cluster validity measures $F(U,C)$ are shown in Fig. 3 and their distributions are listed in Table 2. Since the values are lower than for CNR = 2, the noise sensitivity of fuzzy clustering is obvious. But it is also obvious that the MFCA is clearly less sensitive to noise. Except for the options m_0, n_0 and n_S, which were problematic to the algorithm at CNR = 2, the MFCA yielded improved solutions. Two cases are worth discussing in more detail.

option 0_0: Surprisingly, the FCA performed better with lower CNR, i.e. it could in all cases discriminate the two background levels. Since the FCA is difficult to understand in all its properties, we could only speculate about the reasons for this. Thus we concentrate instead on the comparison with the MFCA. As expected, the MFCA became increasingly stable with increasing number of levels. With four levels, it converged in almost all cases to an improved solution. We also tested this option with five levels and found that the MFCA was fully stable for the solution with $F(U,C) = 0.534$ (2 background clusters + periodic boxcar signal).

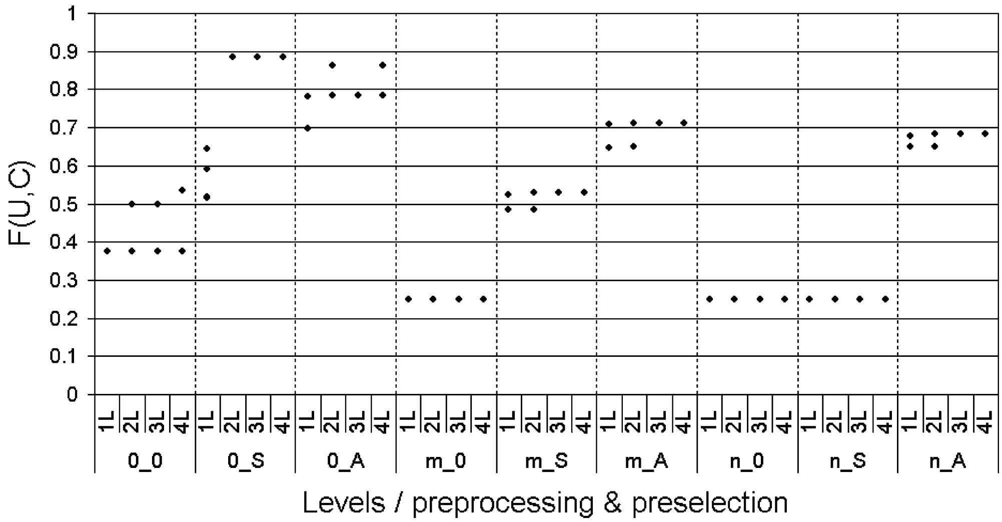


Fig. 3 Cluster validity measures $F(U,C=4)$ for the dataset with CNR = 1. Distributions of the different solutions are listed in Table 2

option 0_A: As was the case with option 0_0, the four levels were not enough to obtain a stable solution. When applying five levels, the ratio was 68% at $F(U,C)=0.78$ vs 32% at $F(U,C)=0.86$. It can be expected that further increasing the number of levels would shift the ratio more towards the solution $F(U,C)=0.68$. But here a limiting factor comes into play: The volumes cannot be downsampled to an arbitrary small size, because the number of time courses left for analysis of a level should be sufficiently large compared with the number of clusters to ensure a meaningful clustering. This is true in particular when the MFCA is combined with some preselection method that further reduces the number of time courses. In our example, the lowest level of the four-level MFCA has 256 time courses, of which 19 survive the autocorrelation criterion. And the five-level MFCA starts with 32 time courses with only five of them surviving the preselection. With six levels, the initial number of time courses is 4, being equal to the number of clusters to classify. It is very likely that the number of time courses will drop below the number of clusters after preselection.

An alternative way to increase the quality of the solution is to choose a different filter for downsampling. A suitable choice is, for example, centred splines [10]. We carried out the same tests for option 0_A with centred splines of order 3 instead of the Haar filter and found that all runs with the three- and four-level MFCA converged to the solution with $F(U,C)=0.86$. Moreover, it took only half the number of iterations needed in the case of the Haar filter. But the filters associated with centred splines have noticeably more coefficients and therefore take a rather long time for downsampling the data to lower levels.

Table 2 Distribution of the different solutions under the same preprocessing and preselection conditions for a given depth of the MFCA (number of levels) for CNR = 1

Preprocessing/preselection Number of levels	Cluster validity measures $F(U,C)$ and their relative frequency		
0_0 2L	0.37 (97%)	0.50 (3%)	
0_0 3L	0.37 (77%)	0.50 (23%)	
0_0 4L	0.37 (3%)	0.53 (97%)	
0_S 1L	0.52 (6%)	0.59 (47%)	0.65 (47%)
0_A 1L	0.70 (43%)	0.78 (57%)	
0_A 2L		0.78 (87%)	0.86 (13%)
0_A 4L		0.78 (97%)	0.86 (3%)
m_S 1L	0.48 (60%)	0.52 (40%)	
M_S 2L	0.48 (7%)	0.53 (93%)	
M_A 1L	0.65 (40%)	0.71 (60%)	
M_A 2L	0.65 (17%)	0.71 (83%)	
N_A 1L	0.65 (73%)	0.68 (27%)	
N_A 2L	0.65 (50%)	0.68 (50%)	

Computational costs

When analysing the noise sensitivity, we also counted the number of iterations needed for each run. From this, we calculated the mean speed-up factors for every preprocessing option as: (mean of number of iterations FCA)/(mean of number of iterations MFCA). Table 3 summarizes the results. In general, the more levels the MFCA has, the faster it is. This follows from the fact that each level of low resolution initializes the preceding level with a good starting point. Whenever the speed-up gets lower with additional levels, the quality of the final

Table 3 Mean speed-up factors of the MFCA for different preprocessing/preselection options

Preprocessing/preselection	CNR = 1			CNR = 2		
	2L	3L	4L	2L	3L	4L
0_0	12	17.5	15.7	20.6	31.0	25.4
0_S	9.8	10.5	8.3	1.8	1.8	1.9
0_A	4.9	4.7	4.8	8.1	9.4	9.5
m_0	3.1	4.2	4.3	3.1	4.2	1.9
m_S	4.9	2.7	2.7	2.8	2.9	2.9
m_A	4.1	4.2	4.3	5.7	8.7	8.8
n_0	3.1	4.2	4.3	3.1	4.1	2.4
n_S	3.1	4.2	4.3	3.1	4.1	2.8
n_A	5.1	5.3	5.3	2.9	4.0	4.0

solution is increased. This can be seen by comparing Table 3 with Fig. 2 and Fig. 3.

Tests with real fMRI data

The cluster validity measures resulting from tests with real fMRI data are shown in Fig. 4. For the case 0 (no preselection, i.e. the full dataset including the background voxels), the FCA and the MFCA up to three levels reached a value of 0.25, meaning that they were not able to separate the dataset into different classes. The four-level MFCA improved the situation only a little. It could identify two clusters (one head-movement-related "signal" and noise) and reached a cluster validity of 0.35. Except for the two cases "MS" and "MSA", the MFCA converged towards solutions superior to the solutions found by the FCA, being more robust the more levels were applied.

The differences between the cluster validity measures of the FCA and the MFCA do not seem to be very large. But it turns out that they are critical for detecting the virtual signal of interest. Consider, for example, the case "SA" with the least difference for $F(U,C)$. The solutions found by the FCA have a validity measure $F(U,C)=0.69$ and the solutions of the three- and four-level MFCA have $F(U,C)=0.75$. While the solution found by the MFCA shows the expected 4 clusters, the FCA failed to find the task-related signal and instead divided one of the head motion clusters into two (see Fig. 5 and Fig. 6). These two clusters have more overlap than the motion cluster has with the task cluster and, accordingly, the cluster validity measure is smaller.

The achieved speed-up of the MFCA is summarized in Table 4. Again, a drop in the speed-up factors corresponds to improved solutions.

Discussion

The MFCA is a significant improvement on the FCA. It is considerably more robust and faster. The results suggest that the more levels are applied in the MFCA, the better the algorithm works. Although this is true in general, there are some limiting factors to be considered. First, the more levels the algorithm has, the more overhead is created for building the multi-resolution pyramid and preprocessing of the data at each level. Second, the number of time courses surviving a preselection should always be checked before processing a low-resolution level. It might happen, in particular at low resolutions, that the number of clusters is almost the same or even exceeds the number of time courses. This

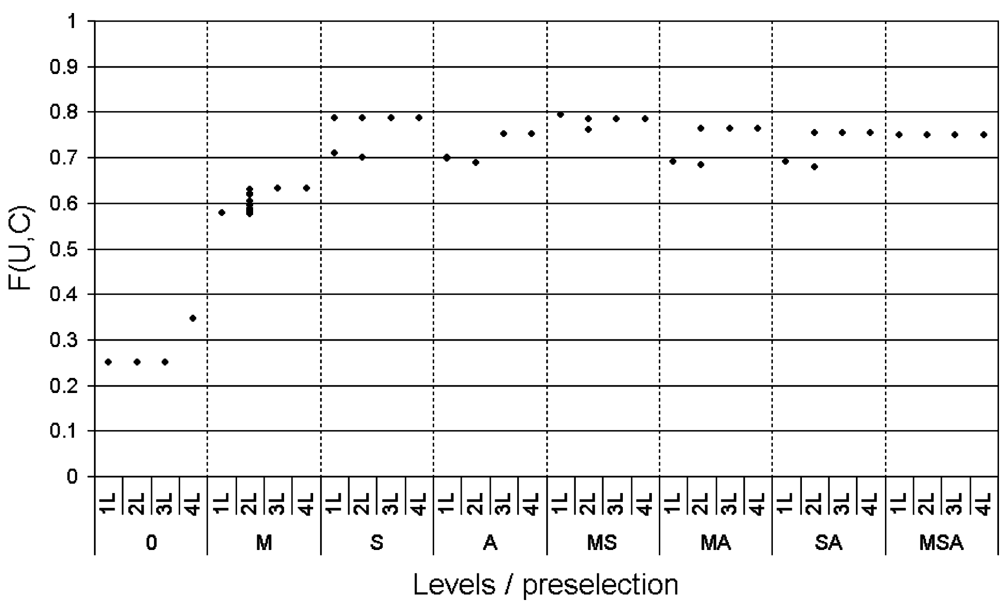


Fig. 4 Cluster validity measures $F(U,C=4)$ for the real fMRI dataset

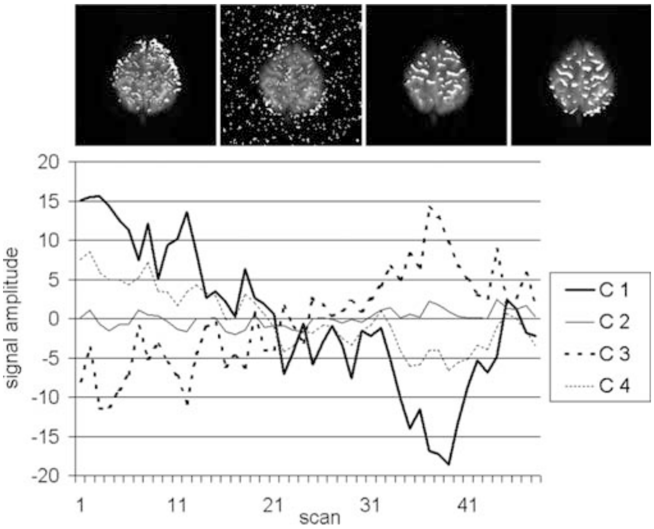


Fig. 5 Clusters and the centroids for the fMRI dataset found by the FCA. Clusters 1 and 4 together with cluster 3 contain the signals associated with head motion, while cluster 2 is the noise. The task-related signal of interest is lost. The cluster validity measure $F(U,C) = 0.69$

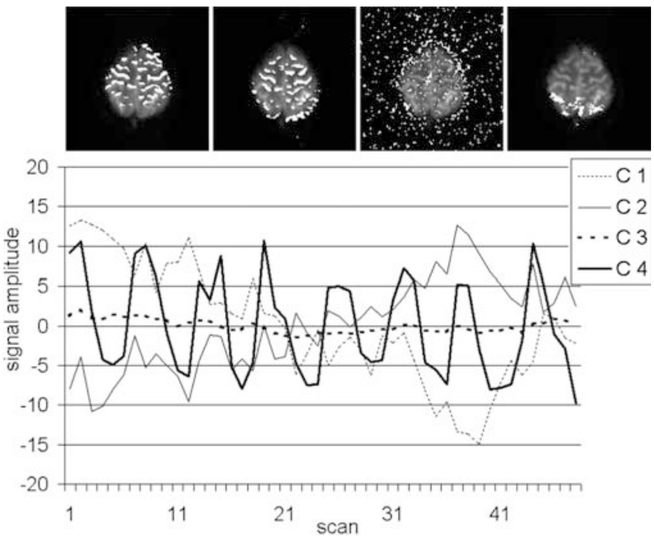


Fig. 6 Clusters and centroids for the fMRI dataset found by the three-level MFCA. Clusters 1 and 2 contain the head motion, cluster 3 is the noise and cluster 4 contains the task-related signal. The cluster validity measure is $F(U,C) = 0.75$

can easily lead to a situation where two clusters have maximum overlap, i.e. their centroids will be equal. If the next level of resolution is then initialized with some centroids being equal, they will not be separated by the FCA at this level. To verify this, suppose that two centroids are equal, i.e. $\mathbf{V}_1 = \mathbf{V}_2$. Then, updating the

Table 4 Mean speed-up factors for the real fMRI dataset

Preselection	2L	3L	4L
0	2.9	3.8	0.6
M	1.1	2.4	2.5
S	2.7	1.9	1.9
A	2.9	1.5	1.8
MS	2.6	1.8	2.6
MA	2.9	3.9	4.0
SA	2.9	2.5	2.6
MSA	3.9	4.2	4.2

membership values with formula (2) yields $u_{i1} = u_{i2}$ for $i = 1,...N$, which in turn again results in $\mathbf{V}_1 = \mathbf{V}_2$ when updating the centroids with formula (3). As a result, the clustering will certainly not be optimal.

At every level it is important to test the number of voxels versus the number of clusters, to use filters that are not too large (with many coefficients) and to perform an MFCA with little depth in the presence of bad results [i.e. deep $F(U,C)$].

The multiresolution technique replaces the necessary random initialization of the FCA by a clustering of low-resolution copies of the dataset. The final FCA applied to the full resolution level becomes very fast and more robust, because it has a good starting point and its convergence is therefore straight and fast.

Designs of fMRI studies of the human brain have become increasingly sophisticated, leading to intricate setups for conventional statistical evaluation techniques to fetch all activation patterns hidden in the data. Therefore, an unsupervised exploration of the data is more suitable. Our results have shown that the fuzzy clustering algorithm is able to reliably find activation patterns in fMRI data. With the MFCA it is even possible to evaluate a complex design including several components that occur as single events in random order in a single analysis [5]. This shows the growing relevance of unsupervised analysis strategies for fMRI. However, as with all these unsupervised strategies, the question remains whether the activation observed corresponds to a task. This remains the focus of future investigation, as would the comparison of the MFCA method with other postprocessing strategies. A p value has been determined for the FCA and can easily be implemented [9]. Potential applications also include the investigation of brain activity and activation during sleep, since in this situation there is no known input factor or applicable box-shaped pattern of activity [11]. Another application could be the measurement of diffusion changes occurring randomly in the penumbral areas of acute stroke [12, 13].

References

1. Friston K, Holmes AP, Worsley KJ, Pioline J-B, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapping 2: 189–210
2. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York
3. Scarth G, McIntyre M, Wowk B, Somorjai R (1995). Detection of novelty in functional images using fuzzy clustering. Proceedings of the 3rd Annual Meeting ISMRM, p 238
4. Fadili MJ, Ruan S, Bloyet D, Mazoyer B (2000) A multistep unsupervised fuzzy clustering analysis of fMRI time series. Hum Brain Mapping 10: 160–178
5. Buerki M, Oswald H, Lovblad KO, Schroth G, Gutbrod K, Schnider A (2002) Evaluation of complex fMRI designs with fuzzy clustering analysis. Psychiatric Res Neuroimaging 114:173
6. Somorjai R, Jarmasz M, Baumgartner R (2000) EVIDENT: a two-stage strategy for the exploratory analysis of functional MRI data by fuzzy clustering. National Research Council, Canada. Available at: http://zeno.ibd.nrc.ca/informatics/pubs/2stage-strategy.pdf
7. Buerki M, Oswald H, Lovblad KO, Schroth G (2001) Additional speed up technique to fuzzy clustering using a multi-resolution approach. Proc SPIE 2001, pp 1141–1150
8. Buerki M, Kiefer C, Oswald H, Lovblad KO, Schroth G (2001) Ein Vergleich verschiedener Multiresolution-Pyramiden für die Fuzzy Cluster basierte Auswertung von fMRI Studien. Presented at the 4th Jahrestagung der Deutschen Sektion der ISMRM, Zurich, Switzerland, 2001
9. Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P (1998). A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn Reson Med 40: 249–260
10. Brigger P, Müller F, Illgner K, Unser M (1999) Centered pyramids. IEEE Trans Image Processing 8: 1254–1264
11. Lovblad KO, Jakob PM, Bassetti C, Griswold M, Ives JR, Matheson J, Edelman RR, Warach S (1999) Silent functional magnetic resonance demonstrates focal activation in REM sleep. Neurology 53: 2193–2195
12. Lovblad KO, Baird A, Schlaug G, Benfield A, Siewert B, Voetsch B, Connor A, Burzynski C, Edelman RR, Warach S (1997) Ischemic lesion volumes in acute stroke by diffusion-weighted magnetic resonance imaging correlate with clinical outcome. Ann Neurol 42: 164–170
13. Schlaug G, Benfield A, Baird A, Siewert B, Lovblad K, Parker R, Edelman RR, Warach S (1999) The ischemic penumbra of human stroke: using functional MRI parameters to define tissue at risk for infarct progression. Neurology 53: 1528–1537